Name : Dhruv Bompilwar
Roll No: CS2-52
PRN : 202401040200

EDS Theory Activity 1

IPL Dataset : 20 problem statements

Problem Statements

1) Top 5 teams with the most runs scored in the dataset

```
[5]  import pandas as pd
     ipl_data = pd.read_csv("deliveries.csv")
```

```
[10] # Top 5 teams with the most runs scored in the dataset

     top_teams = ipl_data.groupby('batting_team')['total_runs'].sum().nlargest(5)
     print("Top 5 teams with the most runs scored:\n", top_teams)
```

```
Top 5 teams with the most runs scored:
 batting_team
Mumbai Indians                42176
Kolkata Knight Riders         39331
Chennai Super Kings           38629
Royal Challengers Bangalore   37692
Rajasthan Royals              34747
Name: total_runs, dtype: int64
```

2) Player with the most runs scored

```
[11] # Player with the most runs scored

     top_batter = ipl_data.groupby('batter')['batsman_runs'].sum().idxmax()
     print("Player with the most runs scored:", top_batter)
```

```
Player with the most runs scored: V Kohli
```

3) Bowler with the most wickets

```
[12] # Bowler with the most wickets

     top_bowler = ipl_data[ipl_data['is_wicket'] == 1].groupby('bowler').size().idxmax()
     print("Bowler with the most wickets:", top_bowler)
```

```
Bowler with the most wickets: YS Chahal
```

4) Most common mode of dismissal

```
[13] # Most common mode of dismissal

common_dismissal = ipl_data['dismissal_kind'].value_counts().idxmax()
print("Most common mode of dismissal:", common_dismissal)
```

```
Most common mode of dismissal: caught
```

5) Average runs scored in each over across all matches

```
[14] # Average runs scored in each over across all matches

avg_runs_per_over = ipl_data.groupby('over')['total_runs'].mean()
print("Average runs scored in each over:\n", avg_runs_per_over)
```

```
Average runs scored in each over:
 over
0     0.981663
1     1.173528
2     1.316099
3     1.356538
4     1.369100
5     1.373055
6     1.103107
7     1.189948
8     1.242087
9     1.224577
10    1.262623
11    1.287686
12    1.301240
13    1.343645
14    1.393504
15    1.434273
16    1.498778
17    1.587839
18    1.646896
19    1.776855
Name: total_runs, dtype: float64
```

6) Total extras conceded by each team

```
[15] # Total extras conceded by each team

extras_by_team = ipl_data.groupby('bowling_team')['extra_runs'].sum()
print("Total extras conceded by each team:\n", extras_by_team)
```

```
Total extras conceded by each team:
 bowling_team
Chennai Super Kings            1842
Deccan Chargers                 659
Delhi Capitals                  725
Delhi Daredevils               1268
Gujarat Lions                   188
Gujarat Titans                  295
Kings XI Punjab                1571
Kochi Tuskers Kerala            110
Kolkata Knight Riders          1957
Lucknow Super Giants            386
Mumbai Indians                 2295
Pune Warriors                   335
Punjab Kings                    480
Rajasthan Royals               1917
Rising Pune Supergiant          111
Rising Pune Supergiants         108
Royal Challengers Bangalore    2040
Royal Challengers Bengaluru     151
Sunrisers Hyderabad            1254
Name: extra_runs, dtype: int64
```

7) Number of sixes hit by each player

```
[16] # Number of sixes hit by each player

sixes = ipl_data[ipl_data['batsman_runs'] == 6].groupby('batter').size()
print("Number of sixes hit by each player:\n", sixes)
```

```
Number of sixes hit by each player:
 batter
A Ashish Reddy     15
A Badoni           24
A Choudhary         1
A Flintoff          2
A Manohar          10
                  ...
YK Pathan         161
YV Takawale         3
Yudhvir Singh       2
Yuvraj Singh      149
Z Khan              2
Length: 451, dtype: int64
```

8) Matches with the highest total runs scored

```
[17] # Matches with the highest total runs scored

highest_scoring_matches = ipl_data.groupby('match_id')['total_runs'].sum().nlargest(5)
print("Matches with the highest total runs scored:\n", highest_scoring_matches)
```

```
Matches with the highest total runs scored:
 match_id
1426268    549
1422126    523
1426280    523
1426281    504
419137     469
Name: total_runs, dtype: int64
```

9) Distribution of wicket types across the dataset

```
[18] # Distribution of wicket types across the dataset

wicket_distribution = ipl_data['dismissal_kind'].value_counts()
print("Distribution of wicket types:\n", wicket_distribution)
```

```
Distribution of wicket types:
 dismissal_kind
caught                   8063
bowled                   2212
run out                  1114
lbw                       800
caught and bowled         367
stumped                   358
retired hurt               15
hit wicket                 15
obstructing the field       3
retired out                 3
Name: count, dtype: int64
```

10) Top 5 bowlers conceding the most runs

```
[19] # Top 5 bowlers conceding the most runs

top_conceding_bowlers = ipl_data.groupby('bowler')['total_runs'].sum().nlargest(5)
print("Top 5 bowlers conceding the most runs:\n", top_conceding_bowlers)
```

```
Top 5 bowlers conceding the most runs:
 bowler
R Ashwin     5435
PP Chawla    5179
B Kumar      5051
RA Jadeja    4917
YS Chahal    4681
Name: total_runs, dtype: int64
```

11) Top partnerships (by runs) in a match

```
[20] # Top partnerships (by runs) in a match

partnerships = ipl_data.groupby(['match_id', 'batter', 'non_striker'])['total_runs'].sum().nlargest(5)
print("Top partnerships (by runs):\n", partnerships)
```

```
Top partnerships (by runs):
 match_id  batter          non_striker
1304112    Q de Kock       KL Rahul       141
829795     AB de Villiers  V Kohli        138
980987     AB de Villiers  V Kohli        132
598027     CH Gayle        TM Dilshan     130
548372     CH Gayle        V Kohli        128
Name: total_runs, dtype: int64
```

12) Find the total number of boundaries (fours) hit by each player

```
[22] # Find the total number of boundaries (fours) hit by each player

fours = ipl_data[ipl_data['batsman_runs'] == 4].groupby('batter').size()
print("Total number of boundaries (fours) hit by each player:\n", fours)
```

```
Total number of boundaries (fours) hit by each player:
 batter
A Ashish Reddy     16
A Badoni           46
A Chopra            7
A Choudhary         1
A Flintoff          5
                  ...
YV Takawale        26
Yashpal Singh       5
Yudhvir Singh       1
Yuvraj Singh      218
Z Khan             11
Length: 538, dtype: int64
```

13) Player dismissed most times by a specific bowler

```
[23] # Player dismissed most times by a specific bowler

dismissal_combinations = ipl_data[ipl_data['is_wicket'] == 1].groupby(['bowler', 'player_dismissed']).size().idxmax()
print("Player dismissed most times by a specific bowler:", dismissal_combinations)
```

```
Player dismissed most times by a specific bowler: ('SP Narine', 'RG Sharma')
```

14) Most consistent batsmen (average runs per match)

```
[24] # Most consistent batsmen (average runs per match)

consistent_batsmen = ipl_data.groupby(['batter', 'match_id'])['batsman_runs'].sum().groupby('batter').mean().nlargest(5)
print("Most consistent batsmen (average runs per match):\n", consistent_batsmen)
```

```
Most consistent batsmen (average runs per match):
 batter
Vivrant Sharma     69.000000
DP Conway          42.000000
B Sai Sudharsan    41.360000
KL Rahul           38.434426
LMP Simmons        37.206897
Name: batsman_runs, dtype: float64
```

15) Number of dot balls bowled by each team

```
[25] # Number of dot balls bowled by each team

dot_balls = ipl_data[ipl_data['total_runs'] == 0].groupby('bowling_team').size()
print("Number of dot balls bowled by each team:\n", dot_balls)
```

```
Number of dot balls bowled by each team:
 bowling_team
Chennai Super Kings           10037
Deccan Chargers                3306
Delhi Capitals                 3732
Delhi Daredevils               6521
Gujarat Lions                  1097
Gujarat Titans                 1851
Kings XI Punjab                7686
Kochi Tuskers Kerala            626
Kolkata Knight Riders         10383
Lucknow Super Giants           1776
Mumbai Indians                11065
Pune Warriors                  1929
Punjab Kings                   2228
Rajasthan Royals               9025
Rising Pune Supergiant          698
Rising Pune Supergiants         539
Royal Challengers Bangalore    9829
Royal Challengers Bengaluru     554
Sunrisers Hyderabad            7556
dtype: int64
```

16) Top 5 matches with the highest number of wickets taken

```
[26] # Top 5 matches with the highest number of wickets taken

top_wicket_matches = ipl_data[ipl_data['is_wicket'] == 1].groupby('match_id').size().nlargest(5)
print("Top 5 matches with the highest number of wickets:\n", top_wicket_matches)
```

```
Top 5 matches with the highest number of wickets:
 match_id
1082625    21
419141     20
1082617    20
1136583    20
1426289    20
dtype: int64
```

17) Total number of wides bowled by each team

```
[27] # Total number of wides bowled by each team

wides_by_team = ipl_data[ipl_data['extras_type'] == 'wides'].groupby('bowling_team').size()
print("Total number of wides bowled by each team:\n", wides_by_team)
```

```
Total number of wides bowled by each team:
 bowling_team
Chennai Super Kings            946
Deccan Chargers                280
Delhi Capitals                 387
Delhi Daredevils               550
Gujarat Lions                   91
Gujarat Titans                 149
Kings XI Punjab                720
Kochi Tuskers Kerala            52
Kolkata Knight Riders          848
Lucknow Super Giants           193
Mumbai Indians                1090
Pune Warriors                  120
Punjab Kings                   234
Rajasthan Royals               949
Rising Pune Supergiant          62
Rising Pune Supergiants         56
Royal Challengers Bangalore    951
Royal Challengers Bengaluru     85
Sunrisers Hyderabad            617
dtype: int64
```

18) Players dismissed on the first ball they faced

```
# Players dismissed on the first ball they faced

first_ball_dismissals = ipl_data[(ipl_data['ball'] == 1) & (ipl_data['is_wicket'] == 1)]['player_dismissed'].unique()
print("Players dismissed on the first ball they faced:\n", first_ball_dismissals)
```

```
Players dismissed on the first ball they faced:
 ['RT Ponting' 'DJ Hussey' 'R Dravid' 'AA Noffke' 'Z Khan' 'MS Dhoni'
 'M Rawat' 'SK Warne' 'AC Gilchrist' 'M Kaif' 'VVS Laxman' 'Shahid Afridi'
 'SB Bangar' 'PA Patel' 'Kamran Akmal' 'K Goel' 'RV Uthappa' 'AM Nayar'
 'Shoaib Malik' 'B Chipli' 'JH Kallis' 'LRPL Taylor' 'MK Pandey'
 'KD Karthik' 'DNT Zoysa' 'D Kalyankrishna' 'LR Shukla' 'AB Dinda'
 'Misbah-ul-Haq' 'DPMD Jayawardene' 'Mohammad Hafeez' 'DJ Bravo'
 'AB de Villiers' 'Mohammad Asif' 'JA Morkel' 'SA Asnodkar' 'V Kohli'
 'DT Patil' 'HH Gibbs' 'RG Sharma' 'YK Pathan' 'RA Jadeja' 'Sohail Tanvir'
 'YV Takawale' 'S Vidyut' 'SK Raina' 'A Chopra' 'CL White' 'SP Goswami'
 'RR Sarwan' 'Yuvraj Singh' 'SC Ganguly' 'DB Ravi Teja' 'SB Styris'
 'J Arunkumar' 'TM Dilshan' 'SE Marsh' 'LPC Silva' 'Y Venugopal Rao'
 'DB Das' 'SM Pollock' 'SR Tendulkar' 'DR Smith' 'SD Chitnis' 'JR Hopes'
 'MF Maharoof' 'A Mishra' 'CK Kapugedera' 'NK Patel' 'SR Watson'
 'Kamran Khan' 'RS Bopara' 'BB McCullum' 'R Bishnoi' 'KC Sangakkara'
 'JD Ryder' 'G Gambhir' 'GC Smith' 'P Kumar' 'IK Pathan' 'RJ Quiney'
 'BAW Mendis' 'S Dhawan' 'Harbhajan Singh' 'ML Hayden' 'S Badrinath'
 'MM Patel' 'MV Boucher' 'AA Bilakhia' 'S Sohal' 'ST Jayasuriya' 'M Vijay'
 'JP Duminy' 'SM Katich' 'RR Raje' 'WA Mota' 'R Bhatia' 'A Symonds'
 'AS Raut' 'GJ Bailey' 'BJ Hodge' 'LA Pomersbach' 'AM Rahane' 'MK Tiwary'
 'NV Ojha' 'P Dogra' 'M Muralitharan' 'AD Mascarenhas' 'MS Bisla'
 'SE Bond' 'V Sehwag' 'M Manhas' 'AB Barath' 'PP Chawla' 'MJ Lumb'
 'AA Jhunjhunwala' 'AT Rayudu' 'KA Pollard' 'DA Warner' 'Y Nagar'
 'FY Fazal' 'SJ Srivastava' 'R Sathish' 'CH Gayle' 'AD Mathews'
 'R Vinay Kumar' 'MD Mishra' 'R Ashwin' 'TL Suman' 'MEK Hussey' 'L Balaji'
 'R McLaren' 'SS Tiwary' 'STR Binny' 'Harmeet Singh' 'Anirudh Singh'
 'RJ Harris' 'IR Jaggi' 'UBT Chand' 'MA Agarwal' 'S Anirudha' 'R Ninan'
 'AJ Finch' 'AL Menaria' 'Sunny Singh' 'DJ Jacobs' 'RV Gomez' 'JE Taylor'
 'KM Jadhav' 'TR Birt' 'SL Malinga' 'TD Paine' 'RR Powar' 'Harpreet Singh'
 'J Botha' 'M Klinger' 'CA Ingram' 'A Singh' 'AC Blizzard' 'OA Shah'
 'KB Arun Karthik' 'CJ Ferguson' 'JJ van der Wath' 'Shakib Al Hasan'
 'F du Plessis' 'MS Gony' 'Ankit Sharma' 'PC Valthaty' 'DT Christian'
 'RE Levi' 'Mandeep Singh' 'KP Pietersen' 'JEC Franklin' 'AB McDonald'
 'MJ Clarke' 'DJ Harris' 'AP Majumdar' 'SPD Smith' 'RP Singh'
```

19) Top 5 fielders with the most catches

```
[29] # Top 5 fielders with the most catches

top_fielders = ipl_data[ipl_data['dismissal_kind'] == 'caught'].groupby('fielder').size().nlargest(5)
print("Top 5 fielders with the most catches:\n", top_fielders)
```

```
Top 5 fielders with the most catches:
 fielder
MS Dhoni          152
KD Karthik        145
AB de Villiers    120
V Kohli           114
SK Raina          106
dtype: int64
```

20) Number of maiden overs bowled by each bowler

```
[30] # Number of maiden overs bowled by each bowler

maiden_overs = ipl_data.groupby(['bowler', 'over'])['total_runs'].sum()
maiden_overs = maiden_overs[maiden_overs == 0].reset_index().groupby('bowler').size()
print("Number of maiden overs bowled by each bowler:\n", maiden_overs)
```

```
Number of maiden overs bowled by each bowler:
 bowler
A Badoni        1
A Choudhary     1
AC Gilchrist    1
Akash Deep      1
D du Preez      1
IS Sodhi        1
M Theekshana    1
MP Yadav        1
N Rana          1
dtype: int64
```